

2022年9月16日

国立研究開発法人理化学研究所

横浜事業所 契約担当役

研究支援部長 富田 久行

(公印省略)

「創薬DXのためのリード化合物創出ワークフロー構築」に関する問合せ事項及び回答

No.	項目	質問内容	回答
1	仕様書 3-1. ワークフローの実装の全般	「ワークフロー及び各アプリは、監督員が指定する統一環境(コンパイラなど処理系、python、python ライブラリ、機械学習系ライブラリなどが統一された環境)に実装すること」となっているが、実装において各アプリおよび導入手順書はご提示頂ける想定でよろしいでしょうか。 また、実装上発生した不具合は開発者にて改修が行われる想定で良いでしょうか。	仕様書に「アプリ(富岳以外で動作確認済み)を提供するのでワークフローへ統合すること。」とある通り、富岳以外で動作確認済みのアプリ(スクリプト)を提供する。ただし、一部のアプリは、統一環境への導入手順書を提供可能である。ワークフローへの統合については基本的に受託者が実施すること。不具合が、本研究所が提供するアプリ固有の原因である場合に限り本研究所が改修するが、それ以外は当然、受託者の責任のもと改修をして頂く。
2	仕様書 3-1. ワークフローの実装の全般	ワークフロー並びにGUI実装はKNIMEを利用した実装を検討しておりますが、よろしいでしょうか。	ワークフロー並びにGUI実装は、受注業者の提案を勘案し、本研究所が決定する。
3	仕様書 3-1. ワークフローの実装の全般	KNIME利用を想定されている場合、KNIME Serverは既どこかの環境に実装(インストール)済でしょうか。実装済である場合実装環境(理研内のオンプレミスサーバや外部クラウド環境等)についてご教示ください。	この場合、ワークフローやアプリの動作に必要なサーバー構築は受託者が行う。
4	仕様書 3-1. ワークフローの実装の全般	ワークフロー並びにGUIをKNIMEではなくスクラッチにて実装する場合、実装想定環境(理研内のオンプレミスサーバや外部クラウド環境等)があればご教示ください。	この場合は、富岳ログインノードまたは本研究所の提供する仮想サーバーを想定している。

No.	項目	質問内容	回答
5	仕様書 3-1. ワークフローの実装の全般	35個程度のアプリ、に関して、具体的なアプリ名を列挙いただけますでしょうか。 また、可能であればアプリ毎に実行形式 (bashなど)、プログラミング言語 (C/C++など) 等も教えてください。	仕様書にある『「アプリ(または計算フロー、以下「計算フロー」は省略)』』について一部のアプリの具体名を列挙する: MP-CAFEE計算フロー (Brown et al., Mol. Inf. 2014, 33, 732, bashなど)、kGCN (https://github.com/clinfo/kGCN , python)、ChemTSv2 (https://github.com/molecule-generator-collection/ChemTSv2 , python)、Replica exchange MDフロー (bashなど)、Autodock vina計算フロー (bashなど)
6	仕様書 3-1. ワークフローの実装の全般	本ワークフローで取り扱うファイル入出力に関して、現状で想定のあるファイル形式を教えてください。ファイル変換は、何パターンほどを想定されているのでしょうか。	XYZ, XTC, TRR, GRO, MOL2, SDF, RDF, CIF, FASTA, CIF, NPZ, PT, PDB, SDF, GRO, PDBQT, XTC, STR, SMILESなどを現状で想定している。ファイル変換は現状で、10パターンほどを想定している。
7	仕様書 3-1. ワークフローの実装の全般	ワークフローの「典型的な実施フロー」について、すでに決まっているものがあれば(実行される具体的なアプリ名と実行回数など)、具体例をいくつか列挙していただくことは可能でしょうか。	ワークフローは仕様書記載内容から一般的な創薬計算フローとして考えられるものを想定している。
8	仕様書 3-1. ワークフローの実装の全般	サーバーをまたがって実行されるワークフローも出てくる可能性があり、また、アプリによっては並列計算が適用されず。通信速度や並列設定を考慮した速度チューニングは仕様範囲外と考えてよいでしょうか。	創薬DX Platformを実行するうえで現実的な計算時間の範囲にするための速度チューニングは仕様を含む。
9	仕様書 3-1. ワークフローの実装の全般	「ワークフローのGUI 上で、アプリによって出力された結果(グラフやテーブルなど)を確認できるように設計・構築すること。」との記載ですが、表示したい結果(グラフやテーブルなど)の想定数をご教示ください。	例えば、MP-CAFEE計算フローについては5程度を想定している。

No.	項目	質問内容	回答
10	仕様書 3-1. ワークフローの実装の全般	<p>「これらのアプリ(富岳以外で動作確認済み)を提供する」「個々のアプリは、複数の実行コマンドを連ねたbash スクリプト、Jupyter notebook、python スクリプト等で記述されている」「OSS も含めてアプリに含まれる実行コマンドや独自プログラムは、単体では富岳で動作するものを、監督員が指定または実装方法を提供」という記述について、以下3点確認させていただきます。</p> <p>①「富岳」で動作確認できていないのは、スクリプト(bash、Jupyter notebook、python等)であり、スクリプトから呼び出されている実行コマンドやプログラムは、今回の対象のものは、単体ではすべて「富岳」で稼働確認済みであるという認識でよいでしょうか？</p> <p>② アプリのスクリプトについても、基本的には「富岳」で実行可能な状態で提供いただけると考えてよろしいでしょうか？(「富岳」のパラメータ等を設定したスクリプトを作成し、そこからアプリのスクリプトを呼び出す形で実行できることを想定)</p> <p>③ Jupyter notebookで記述されているアプリがあると記載されていますが、Jupyter notebookをWebブラウザ上ではなく、コマンドラインから実行できるような形で提供され、実行環境や実行手順については、「富岳」で実行する他のOSS同様に、ご提供いただけると考えてよいでしょうか？</p>	<p>①について： ほぼすべての「スクリプトから呼び出されている実行コマンドやプログラム」は富岳で動作確認済みである。富岳で動作しないものは、別のサーバーで実行を指示する場合がある。</p> <p>②について： アプリのスクリプトについては、仕様にある通り「アプリ(富岳以外で動作確認済み)を提供する」。</p> <p>③について： 受託者が創薬DX PlatformIに実装するうえで必要な修正を行うこと。</p>
11	仕様書 3-1. ワークフローの実装の全般	<p>「本件で構築した創薬DXPF を、富岳とは別のクラウドに移植が可能であること。その場合の、具体的な移植方針を報告書にまとめ提出すること」と記載されていますが、報告書に含める移植方針の範囲は、本仕様書で実装対象となっているワークフロー部分のみ(理研から指定、実装方法を提供いただく「OSSも含めてアプリに含まれる実行コマンドや独自プログラム」は対象外)と考えてよろしいでしょうか？</p>	<p>本仕様書で実装対象となっているワークフロー部分を対象範囲とする。ただし、本研究所から指定、実装方法を提供する「OSSも含めてアプリに含まれる実行コマンドや独自プログラム」が、報告書に記載されたワークフローに対して、どの部分に該当するかを明示すること。</p>
12	仕様書 3-1. ワークフローの実装の全般	<p>「本件で構築した創薬DXPF を、富岳とは別のクラウドに移植が可能であること。その場合の、具体的な移植方針を報告書にまとめ提出すること」と記載されていますが、「別のクラウド」とは、富岳やFOCUSなどのようなスパコン(SSH接続 + ジョブ投入ができるシステム)と考えてよいでしょうか。可能性のあるクラウドサービスや計算機のアーキテクチャ(x86やarm64など)を教えてください。</p>	<p>ISMAPに記載されているクラウドやHPCIのスーパーコンピュータを想定している。</p>

No.	項目	質問内容	回答
13	仕様書 3-1. ワークフローの実装の全般	<p>「アプリによっては富岳外のサーバーで、リモート実行を行う場合がある」と記載されていますが、以下4点確認させてください。</p> <p>①「富岳」外のサーバーが何種類あり、それぞれのサーバーが設置されている場所をご教示いただけないでしょうか？</p> <p>②「富岳」外で実行する具体的なアプリ名をご教示いただけないでしょうか？</p> <p>③「富岳」外のサーバーで実行するアプリについては、リモート実行は稼働確認済みで、実行方法についてはご提供いただけたらと考えてよいでしょうか？</p> <p>④「富岳」で実行するアプリから、直接「富岳」外のサーバーにあるアプリを呼び出すようなケースもあるのでしょうか？</p>	<p>①②④について:本研究所の指示のもと決定する。なお、本研究所外に設置されたDBにクエリを発行することがある。</p> <p>③について:リモート実行は稼働確認済みではない。受注業者が設計の上実装すること。</p>
14	仕様書 3-2. 「化合物創出」ワークフロー	<p>「各化合物についての計算済みの情報はどのモジュールからでも読み出せるように設計・構築すること」と記載されていますが、アプリの実行結果は、各アプリからアクセス可能なストレージに保存し、他のアプリは、必要に応じてディレクトリやファイルを指定して読み出すことを想定しています。データベース等を利用して、出力データや過去データの履歴管理を行うことは想定していませんが、問題ないでしょうか？</p>	<p>『「各化合物についての計算済みの情報はどのモジュールからでも読み出せるように設計・構築すること」と記載されていますが、アプリの実行結果は、各アプリからアクセス可能なストレージに保存し、他のアプリは、必要に応じてディレクトリやファイルを指定して読み出すことを想定しています』について、認識の通りである。</p> <p>データベース等を利用して、出力データや過去データの履歴管理を行うかは、受注業者が提案する設計の上、本研究所が決定する。</p>
15	仕様書 3-3. リグレーションテスト設計と必要計算時間の見積もり機能	<p>「テスト内容は、各モジュールに対して3パターン程度ずつ、10パターン程度のテストを構築する」と記載されていますが、以下2点について確認させてください。</p> <p>① 5つのモジュールに3パターン程度ある場合、合計15パターン程度になりますが、実際には、モジュールによってはそれよりも少ないケースでよいため、合計で10パターン程度のテストと考えるとよろしいでしょうか？</p> <p>②リグレーションテスト用のワークフローは、仕様書3-1で実装するワークフロー(複数のモジュールにまたがるワークフローを想定)とは、別のワークフローを作成するという認識でよろしいでしょうか？</p>	<p>①について:認識のとおりである。</p> <p>②について:構築した創薬DX Platformのワークフローの一部を利用するかたちでテストを作成すること。</p>

No.	項目	質問内容	回答
16	仕様書 3-1. ワークフローの実装の全般	「どちらのワークフローも複数の「モジュール」から構成されている」に関して。 「モジュール」というのは、ワークフローを実装する上で、何か考慮が必要な概念になりますでしょうか。単にアプリの集合体、というだけの位置づけであり、実際に呼び出して利用するのは「アプリ」でしょうか。	認識のとおりである。開発・テスト・設計書をモジュール単位に実行することを想定しているが、実際の作業工程は受託者の提出する作業工程表に基づき、本研究所が決定する。
17	仕様書 3-1. ワークフローの実装の全般	「ワークフローの実施フロー（使用するアプリの選択・順番・回数）をユーザーが設定できる仕組みを設計・実装すること。同時に、監督員が典型的な実施フローを指定するので、この情報を基にして、10 通り程度の実施フローのプリセットを用意すること。」とありますが、用意するのはGUI、CUIともに10程度でしょうか。それともGUI、CUI合わせて10個程度でしょうか。	プリセットはCUI、GUI合わせて10個を上限とする。ただし、ワークフローの実施フロー（使用するアプリの選択・順番・回数）をユーザーが設定できる仕組みも設計・実装すること。
18	仕様書 3-1. ワークフローの実装の全般	「ワークフローの概要に関する説明スライドを提出すること。ワークフロー全体のプログラム、ユーザーが利用することを想定したワークフローの日本語のマニュアルを提出すること。」とありますが、 「ワークフロー全体のプログラム」とはどの作成物を意味しているかご教示いただけますでしょうか。 KNIMEの場合作成するワークフローの個数分（GUI10通りなら10個）のマニュアルでしょうか。	「ワークフロー全体のプログラム、ユーザーが利用することを想定したワークフローの日本語のマニュアル」は、アプリが統合されたプラットフォームの全般的な説明と、プラットフォーム上での各モジュール・アプリの利用法を合わせたものをいう。加えて、プリセットの個数分の実行手順の説明を含むこと。
19	仕様書 3-1. ワークフローの実装の全般	「ワークフローの概要に関する説明スライドを提出すること。ワークフロー全体のプログラム、ユーザーが利用することを想定したワークフローの日本語のマニュアルを提出すること。」とありますが、 ・ワークフローの概要に関する説明スライド → パワーポイントとで作った概要説明 ・ワークフロー全体のプログラム ・ユーザーが利用することを想定したワークフローの日本語のマニュアル → 利用者マニュアル（システム管理者向けのマニュアルは対象外） の3つが納品対象でよいでしょうか。	「ワークフローの概要に関する説明スライドを提出すること。ワークフロー全体のプログラム、ユーザーが利用することを想定したワークフローの日本語のマニュアルを提出すること。」の項目に関しては以下を納品すること。 ・ワークフローの概要に関する説明スライド（例：パワーポイントで作った概要説明） ・ワークフロー全体のプログラム ・ユーザーが利用することを想定したワークフローの日本語のマニュアル ・システム管理者向けのマニュアル 本項目以外にも仕様書にある情報を提出すること（例えば、ワークフローのファイル入出力設計書・計算フロー設計書）
20	仕様書 6.実施体制	業務遂行にあたっては弊社からの再委託も想定していますが問題ありますでしょうか。	応札時に再委託先やその内容について提出し、本研究所の了承を得ること。

No.	項目	質問内容	回答
21	仕様書 3-1. ワークフローの実装の全般	「どちらのワークフローも複数の「モジュール」から構成されている。個々のモジュールはさらに小さい「アプリ(または計算フロー、以下「計算フロー」は省略)」からなる。これらのアプリ(富岳以外で動作確認済み)を提供するのでワークフローへ統合すること。各アプリのワークフローへの統合にあたって、ワークフロー全体のためのファイル入出力(接続先DB・通信を含む)の設計書を構築すること。このために、必要に応じて開発担当者(理研および共同研究先)へのヒアリングを行うこと。実装するアプリは35個程度を想定している。」とありますが、実装するアプリは35個を上限として考えてよろしいでしょうか。	実装するアプリは35個を上限とする。
22	仕様書 3-1. ワークフローの実装の全般	「ワークフローの実施フロー(使用するアプリの選択・順番・回数)をユーザーが設定できる仕組みを設計・実装すること。同時に、監督員が典型的な実施フローを指定するので、この情報を基にして、10通り程度の実施フローのプリセットを用意すること。」とありますが、実施フローのプリセットは10通りを上限として考えてよろしいでしょうか。	実施フローのプリセットは10通りを上限とする。
23	全般	作業スケジュール作成前に作業期間中のメンテナンス、縮小運転、停電等の計画を事前にご連絡いただくことは可能でしょうか。	計画を提供することは可能である。
24	全般	昨今の世界的情勢の影響で光熱水費高騰対策として計算機の一部を停止した縮小運用や計画外のメンテナンス等が発生したことが原因で、計画通りリソースが使用できない場合、「富岳」を利用しての確認作業に影響があり、作業スケジュールが大幅に遅延することが想定されます。「富岳」の運用が原因で、作業に影響がある場合の対処、対応を教えてくださいませんか。	状況に応じて、双方協議となる。富岳のアーキテクチャと類似したシステムでの実施を指示する場合がある。